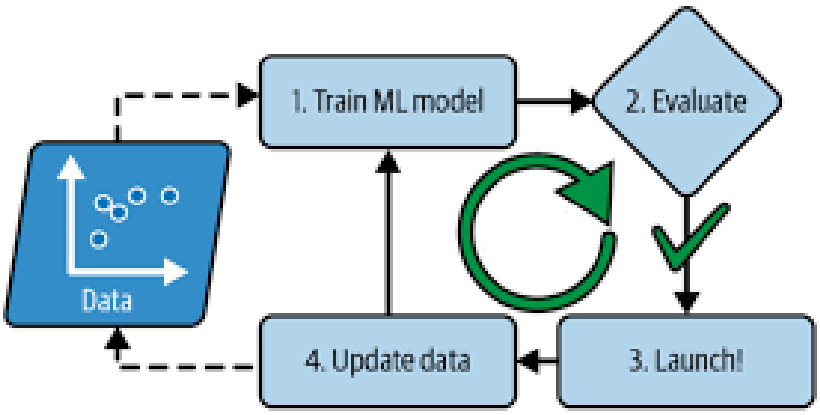
Data Science | 30 Days of Machine Learning | Day - 2

Educator Name: Nishant Dhote
Support Team: **+91-7880-113-112**

What is Batch / Offline Machine Learning?

What is Online Machine Learning?

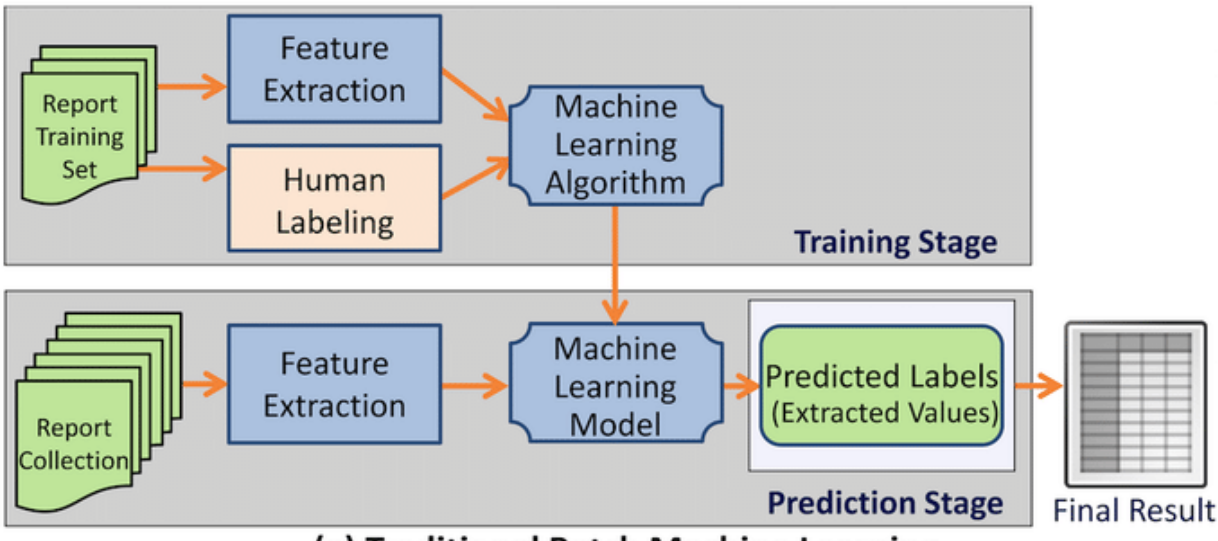
Difference Between Online Vs Offline Machine Learning?



Batch Machine Learning (Offline ML):

In batch learning, data is accumulated over a period of time. The machine learning model is then trained with this accumulated data from time to time in batches. It is the direct opposite of online learning because the model is unable to learn incrementally from a stream of live data. In batch learning, the machine learning algorithm updates its parameters only after consuming batches of new data.

The fact that models are trained with large batches of accumulated data means that more time and resources such as CPU, memory space, and disk input/output are needed.



(a) Traditional Batch Machine Learning

Batch size (machine learning)

Batch size is a term used in machine learning and refers to the number of training examples utilized in one iteration. The batch size can be one of three options:

Batch mode: where the batch size is equal to the total dataset thus making the iteration and epoch values equivalent

Mini-batch mode: where the batch size is greater than one but less than the total dataset size. Usually, a number that can be divided into the total dataset size.

Stochastic mode: where the batch size is equal to one. Therefore, the gradient and the neural network parameters are updated after each sample.

Advantages: (When we use?)

- Lower production costs
- Flexibility
- Better quality control
- Increased efficiency

Disadvantages:

- Large-scale failures
- They are monolithic in nature
- Complex to manage
- Require very expensive

Online Machine Learning:

Online machine learning is a method of machine learning where the model incrementally learns from a stream of data points in real-time. It's a dynamic process that adapts its predictive algorithm over time, allowing the model to change as new data arrives. This method is incredibly significant in today's rapidly evolving data-rich environments because it can provide timely and accurate predictions.

Real time example: Chat Bot | YouTube or OTT Recommended | Stock | Fraud Detection

Library use in Python: River Library (https://riverml.xyz/dev/) | Vowpal Wabbit (https://vowpalwabbit.org/)

Advantages:(What are the Benefits of Online Machine Learning?)

Adaptability. Just like the cyclist learning as they go, online machine learning can adapt to new patterns in the data, improving its performance over time.

Scalability. Since online learning processes data one at a time, it doesn't require the storage capacity that batch learning does. This makes it scalable to big data applications.

Real-time predictions. Unlike batch learning that might be outdated by the time it's implemented, online learning provides real-time insights, which can be critical in many applications like stock trading and health monitoring.

Efficiency. As online machine learning allows for continuous learning and updating of models, this can lead to faster and more cost-efficient decision-making processes.

Disadvantages:(What are the Limitations of Online Machine Learning?)

Sensitive to sequence or Tricky to Use. The order in which the data is presented can impact the learning process. An unusual data point can significantly alter the model's parameters, leading to decreased accuracy.

Less control over training or Risky. Unlike batch learning, where you can control the training process, online learning is always on. An unexpected influx of bad quality data can lead to poor predictions.

Lack of interpretability. Online learning algorithms, especially those based on deep learning or neural networks, can be highly complex and difficult to interpret. This lack of interpretability can make it challenging to understand and explain the model's decisions.

If we facing problem in online learning system so real time technique, we use: We use anomaly detection system so use switch the system in Offline Mode or rollback the system online to offline batch learning.

Difference Between Online Vs Offline Machine Learning?

Offline Learning	Features	Online Learning
Less complex as model is constant	Complexity	Dynamic complexity as the model keeps evolving over time
Fewer computations, single time batch-based training	Computational Power	Continuous data ingestions result in consequent model refinement computations
Easier to implement	Use in Production	Difficult to implement and manage
Image Classification or anything related to Machine Learning - where data patterns remains constant without sudden concept drifts	Applications	Used in finance, economics, heath where new data patterns are constantly emerging
Industry proven tools. E.g. Sci-kit, TensorFlow, Pytorch, Keras, Spark Mlib	Tools	Active research/New project tools: E.g. MOA, SAMOA, scikit-multiflow, streamDM

Instance Based Machine Learning

Model Based Machine Learning

Instance Based Vs Model Based Machine Learning

Think about how human learn, answer is memorised or generalized

Learning → Memorising (Example of Instanced Based ML)

Learning → Generalized (Example of Model Based ML)

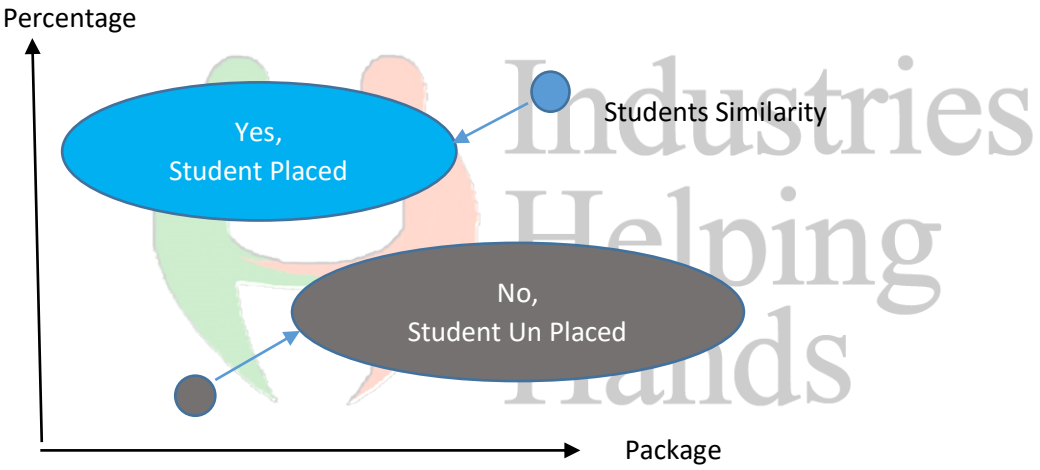
 # Instance Based Machine Learning:

The Machine Learning systems which are categorized as instance-based learning are the systems that learn the training examples by heart and then generalizes to new instances based on some similarity measure. It is called instance-based because it builds the hypotheses from the training instances. It is also known as memory-based learning or lazy-learning (because they delay processing until a new instance must be classified).

The time complexity of this algorithm depends upon the size of training data. Each time whenever a new query is encountered, its previously stores data is examined. And assign to a target function value for the new instance.

Example: Sample Dataset

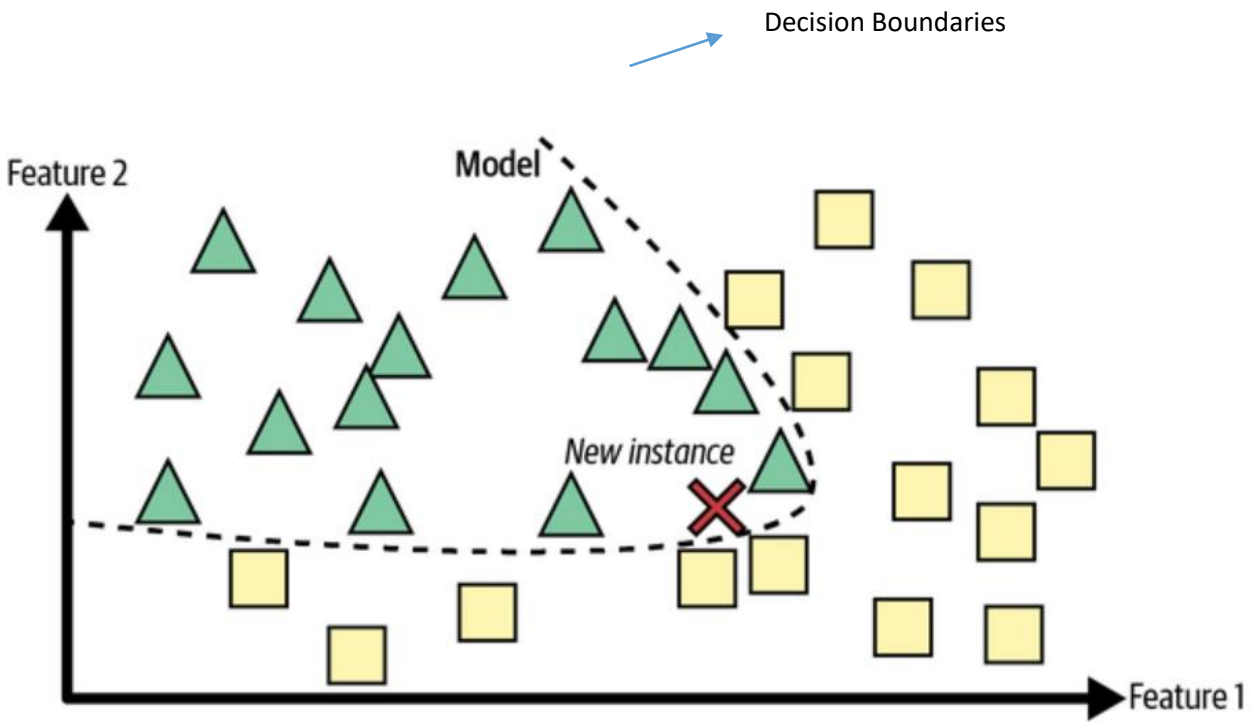
S.No	Student_Name	St_Percentage	St_CGPA	St_Placement	Package
1	Nishant	87%	7.8	Yes	8 LPA
2	Swati	75%	7.2	No	0 LPA
3	Ragini	91%	8.4	Yes	10 LPA
4	Salman	68%	6.5	No	0 LPA
.					
.					
10K	XYZ	00%	0.0	No	0



The new instance would be classified as a triangle because the majority (2 Triangles vs 1 Square) of the most similar instances belong to that class

 # Model Based Machine Learning:

Model-based learning (also known as structure-based or eager learning) takes a different approach by constructing models from the training data that can generalize better than instance-based methods. This involves using algorithms like linear regression, logistic regression, random forest, etc. trees to create an underlying model from which predictions can be made for new data points. The picture below represents how the prediction about the class is decided based on boundary learned from training data rather than comparing with learned data set based on similarity measures.



The model based learning approach has several benefits over instance-based methods, such as faster processing speeds and better generalization capabilities due to its use of an underlying model rather than relying solely on memorized examples. However, this approach requires more time and effort to develop and tune the model for optimal performance on unseen data sets.

 # Instance Based Vs Model Based Machine Learning

Model-Based Machine Learning	Instance-Based Machine Learning
Prepare the data for model training	Prepare the data for model training. No difference here
Train model from training data to estimate model parameters i.e. discover patterns	Do not train model. Pattern discovery postponed until scoring query received
Store the model in suitable form	There is no model to store
Generalize the rules in form of model, even before scoring instance is seen	No generalization before scoring. Only generalize for each scoring instance individually as and when seen
Predict for unseen scoring instance using model	Predict for unseen scoring instance using training data directly
Can throw away input/training data after model training	Input/training data must be kept since each query uses part or full set of training observations
Requires a known model form	May not have explicit model form
Storing models generally requires less storage	Storing training data generally requires more storage

#Challenges in Machine Learning:

#Application of Machine Learning:

#Machine Learning Development Life Cycle:

 # Challenges in Machine Learning:

- Data Collection
- Insufficient Data
- Non Reprehensive Data (Sampling Noise, Sampling Bias)
- Poor Quality Data
- Irrelevant Features
- Overfitting & Under fitting
- Software Integration
- Deployment & Cost Involved

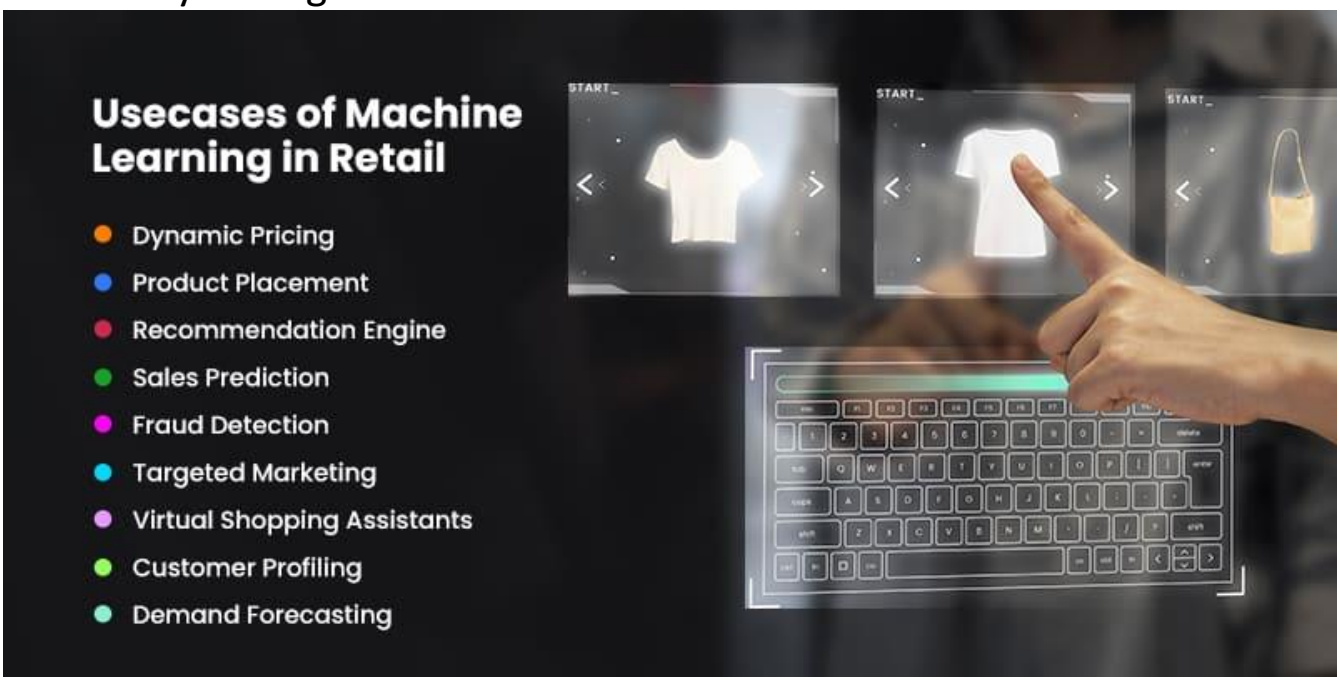
 # Application of Machine Learning:

1. Retail / E-Commerce Sector
2. Banking and Finance Sector
3. Logistics and Transportation Sector
4. Manufacturing Sector
5. Social Networking Sector
6. Many More Sector

 # Machine Learning Development Life Cycle (MLDLC/MLDC):

1. Frame the problem
2. Gathering data
3. Data pre processing
4. Exploratory data analysis
5. Feature engineering & selection
6. Model training, evaluation and selection
7. Model deployment
8. Testing
9. Optimize

Data Story Telling: Curious Data Minds



Case Study Link: https://www.matellio.com/blog/machine-learning-use-cases-in-retail/